COMP90042

Web search and text analysis

Workshop Week 5

Your tutor

- Winn Chow (Senior Tutor)
- winn.chow1@unimelb.edu.au
- Office: Doug McDonell - 9.23
- Here, you can find my workshop slides:
- https://github.com/winnchow/COMP90042-Workshops

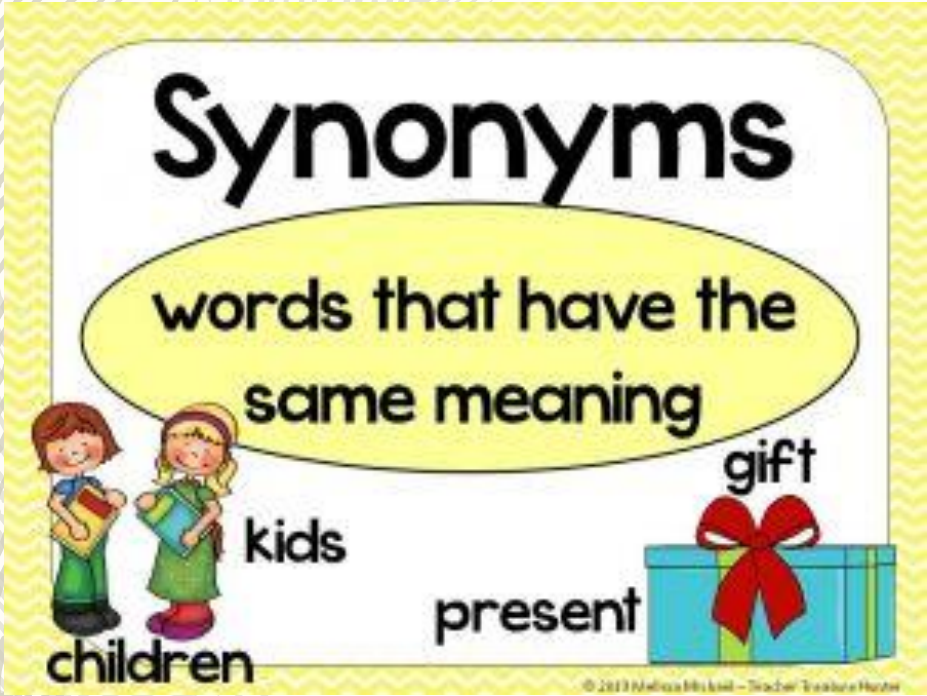
Q1

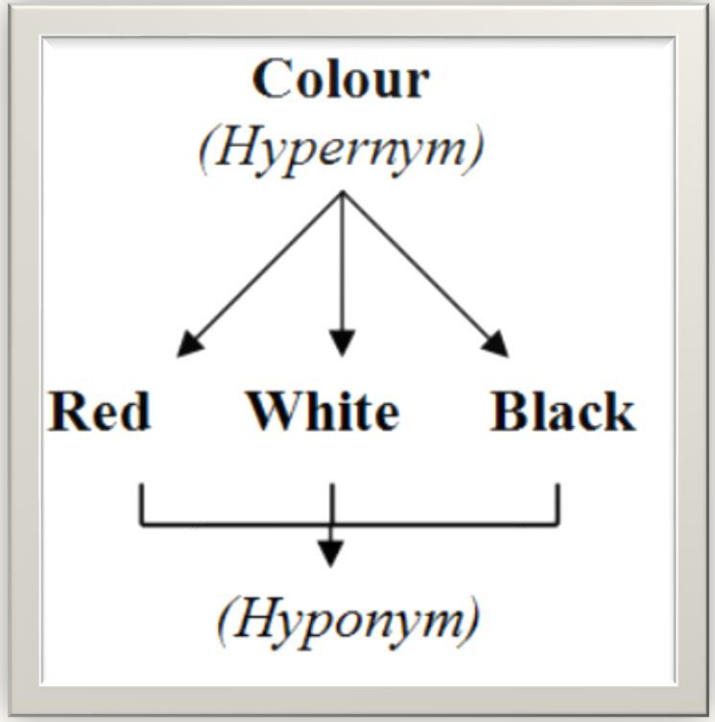
1. Give illustrative examples that show the difference between:
   (a) **Synonyms** and **hypernyms**
   (b) **Hyponyms** and **meronyms**

Q1

- **Meronym:** Part of a whole

- **Holonym:** The whole to which parts belong





Q2

2. Using some Wordnet visualisation tool, for example, `http://wordnetweb.princeton.edu/perl/webwn` and the Wu & Palmer definition of **word similarity**, check whether the word *information* is more similar to the word *retrieval* or the word *science* (choose the sense which minimises the distance). Does this mesh with your intuition?

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: information | Search WordNet

Display Options: (Select option to change) | Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

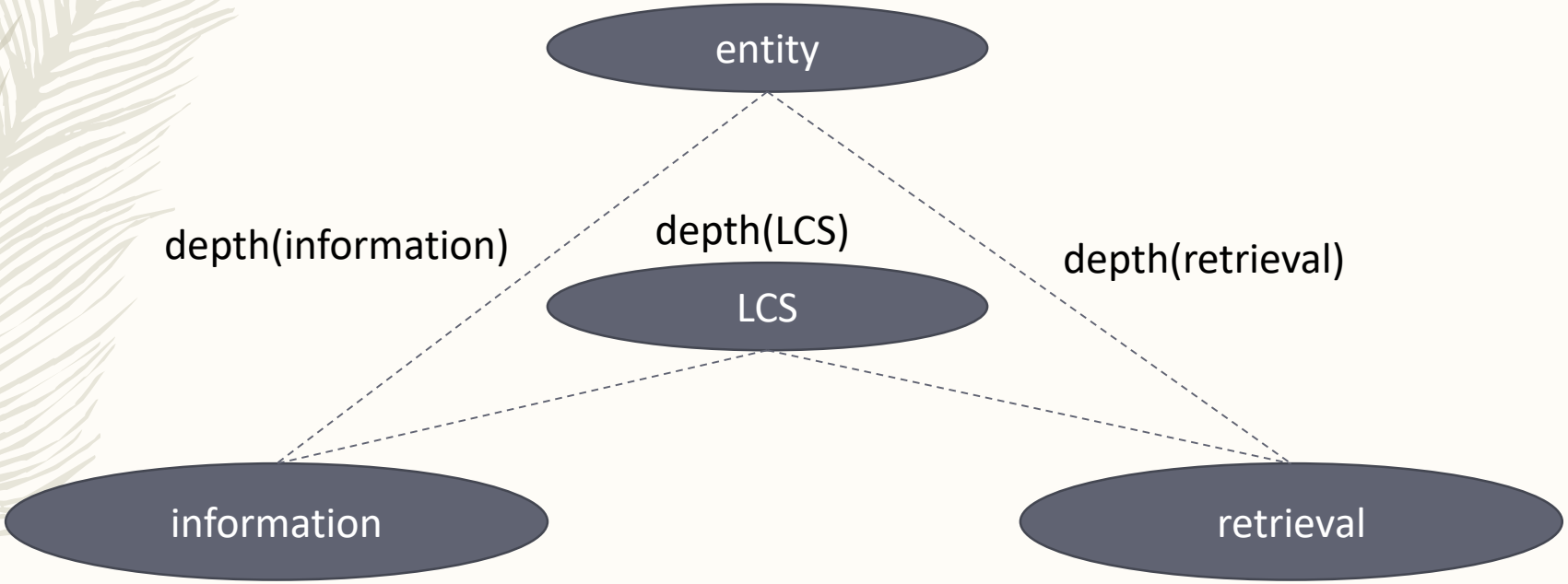
Noun

- S: (n) **information**, info (a message received and understood)
  - *direct hyponym* / *full hyponym*
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) message, content, subject matter, substance (what a communication that is about something is about)
      - S: (n) communication (something that is communicated by or to or between people or groups)
        - S: (n) abstraction, abstract entity (a general concept formed by extracting common features from specific examples)
          - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
  - *derivationally related form*
- S: (n) **information** (knowledge acquired through study or experience or instruction)
- S: (n) **information** (formal accusation of a crime)
- S: (n) data, **information** (a collection of facts from which conclusions may be drawn) *"statistical data"*
- S: (n) **information**, selective information, entropy ((communication theory) a numerical measure of the uncertainty of an outcome) *"the signal contained thousands of bits of information"*

Solution 1: include depth information (Wu & Palmer)

* Use path to find lowest common subsumer (LCS)
* Compare using depths

$$\text{simwup}(c_1,c_2) = \frac{2 * \text{depth}(\text{LCS}(c_1,c_2))}{\text{depth}(c_1) + \text{depth}(c_2)}$$



$$\text{sim}(\text{information}, \text{retrieval}) = \frac{2 \times 1}{5+8} = \frac{2}{13} \approx 0.154$$

$$\text{simwup}(c_1,c_2) = \frac{2 * \text{depth}(\text{LCS}(c_1,c_2))}{\text{depth}(c_1) + \text{depth}(c_2)}$$

depth(information) = 5

			entity		
			abstraction...		
			communication		
			message...		entity
1	entity	entity	statement	entity	abstraction...
2	abstraction...	abstraction...	pleading	abstraction...	measure
3	communication	psychological...	charge...	group...	system of meas...
4	message...	cognition...	accusation...	collection...	information meas...
5			information		

depth(LCS) = 1 depth(LCS) = 2 depth(LCS) = 2

1	entity	entity	
2	physical...	abstraction...	
3	process...	psychological...	entity
4	processing	cognition...	abstraction...
5	data process...	process...	psychological...
6	operation	basic cog...	event
7	computer op...	memory...	act...
8		retrieval	

depth(retrieval) = 8 depth(retrieval) = 8 depth(retrieval) = 6

$$\text{simwup}(c_1,c_2) = \frac{2*\text{depth}(\text{LCS}(c_1,c_2))}{\text{depth}(c_1) + \text{depth}(c_2)}$$

depth(information) = 5

1	entity	entity	entity		entity
2	abstraction...	abstraction...	abstraction...		abstraction...
3	communication	psychological...	communication		measure
4	message...	cognition...	message...		system of meas...
5			statement		information meas...
			pleading		
			charge...		
			accusation...		
			information		

depth(LCS) = 1 depth(LCS) = 4 depth(LCS) = 3

1	entity	entity	
2	physical...	abstraction...	
3	process...	psychological...	entity
4	processing	cognition...	abstraction...
5	data process...	process...	psychological...
6	operation	basic cog...	event
7	computer op...	memory...	act...
8		retrieval	

depth(retrieval) = 8 depth(retrieval) = 8 depth(retrieval) = 6

Wu & Palmer Similarity

		information				
		1	2	3	4	5
	1	0.154	0.154	0.118	0.154	0.143
retrieval	2	0.308	**0.615**	0.235	0.308	0.286
	3	0.364	0.545	0.267	0.364	0.333

Q4

4. For the following term co-occurrence matrix (suitably interpreted):

	cup	not (cup)
world	55	225
not (world)	315	1405

(a) Find the Point-wise Mutual Information (PMI) between these two terms in this collection.

(b) What does the value from (a) tell us about **distributional similarity**?

$$PMI(x,y) = \log_2 \frac{p(x,y)}{p(x)p(y)}$$

Q4a PMI

	cup	not (cup)	Total
world	55	225	**280**
not (world)	315	1405	**1720**
Total	**370**	**1630**	**2000**

$$PMI(x,y) = \log_2 \frac{p(x,y)}{p(x)p(y)}$$

Q4a PMI

$$
\begin{aligned}
P(w) &= 280/2000 = 0.14 \\
P(c) &= 370/2000 = 0.185 \\
P(w,c) &= 55/2000 = 0.0275 \\
PMI(w,c) &= \log_2 \frac{P(w,c)}{P(w)P(c)} \\
&= \log_2 \frac{0.0275}{0.14 \times 0.185} \\
&\approx 0.0865
\end{aligned}
$$

This value is slightly positive, which means that the two events occur together (in documents) slightly more commonly than would occur purely by chance. There is some possibility that `world` and `cup` occurring together is somehow meaningful for documents in this collection.

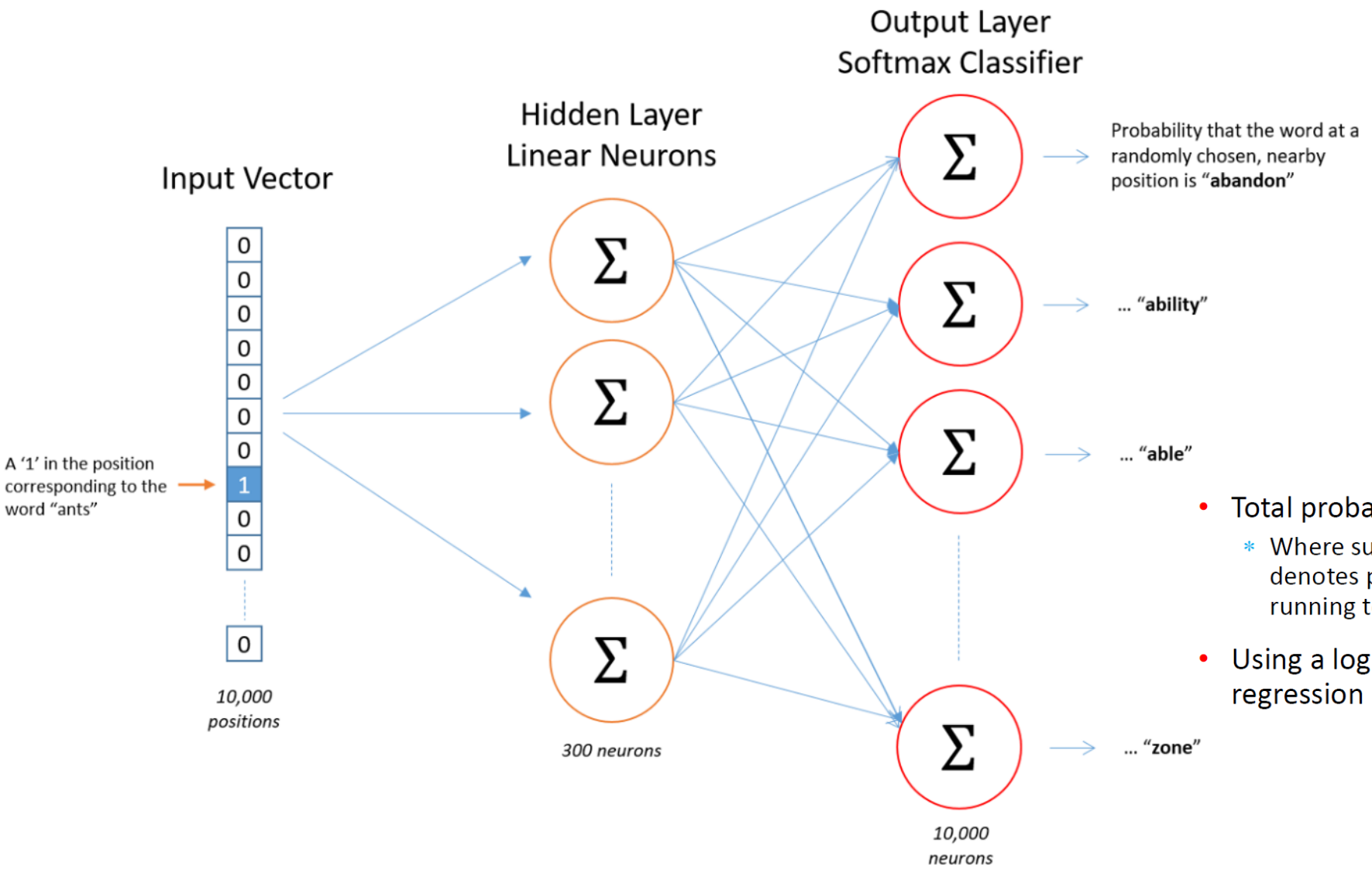
Q6

6. What is a **word embedding** and how does it relate to **distributional similarity**?

   (a) What is the difference between a **skip-gram** model and a **CBOW** model?

   (b) How are the above models trained?

Skip-gram neural network model



- Total probability defined as $\prod_{l \in -L,\ldots,-1,1,\ldots,L} P(w_{t+l}|w_t)$
  - Where subscript denotes position in running text
- Using a logistic regression model $P(w_k|w_j) = \frac{\exp(c_{w_k} \cdot v_{w_j})}{\sum_{w' \in V} \exp(c_{w'} \cdot v_{w_j})}$

https://towardsdatascience.com/word2vec-skip-gram-model-part-1-intuition-78614e4d6e0b

Q6a

- We're going to have a representation of words (based on their contexts) in a **vector space**, such that other words "nearby" in the space are similar
- This is broadly the same what we expect in distributional similarity, e.g. "you shall know a word by the company it keeps."